Constraint Grammars

Thierry Martinez

Acknowledgments to Rémy Haemmerlé
for the original impulse to this work.

Contraintes Project–Team, INRIA Paris–Rocquencourt, France

21 July 2008

Outline

Constraint Sets.

Observational confluence

```
nil(L) \ positive_list(L) ⟺ true.
list(L, H, T) \ positive_list(L) ⟺
        H > 0, positive_list(T).
```

- Confluent for queries modelling a list (first arguments of `list` tokens are pair-wise distincts).
- Non-confluent for queries modelling a tree (with two `list` tokens sharing a common first argument).

Characterizing (L)CC Stores

Describing Global Constraints

Given a set of constraints S:

- either there exists a satisfied constraint in S:

  $\text{REGULAR}(x_1, \ldots, x_n, L)$ with $S = \{x_1 = u_1 \wedge \cdots \wedge x_n = u_n\}_{u \in L}$

- or all the constraints of S are satisfied:

  $\text{ALL-DIFFERENT}(x_1, \ldots, x_n)$ with $S = \left\{x_i \neq x_j\right\}_{i \neq j}$

Linear Constraint Systems.

Let V be a set of variables.

- If C is a set of formulae over the set of variables V, let $\mathfrak{C}[C]$ be the closure of C by conjunction ($\otimes$) and variable hiding ($\exists$).
- If $\Vdash$ is a binary relation over C, let $\widehat{\Vdash}$ be the closure of $\Vdash$ by the rules of linear logic.

Definition

A (linear) **constraint system** is a pair $(C, \vdash)$ where:

- C is a set of formulae closed by conjunction ($\otimes$) and variable hiding ($\exists$);
- $\vdash \subseteq C^2$ is a relation closed by the rules of linear logic.

Examples of Constraint Systems

Let Σ be a signature. Let $t(\Sigma) \doteq \left\{ f(v_1, \ldots, v_n) | f/n \in \Sigma \right\}$.

linear-token system

The **linear-token system** over Σ is the constraint system $\textsc{tok}(\Sigma) \doteq (\mathfrak{C}[t(\Sigma)], \widehat{\varnothing})$.

linear-token system with equality

The **linear-token system with equality** over Σ is the constraint system $\textsc{tok}_=(\Sigma) \doteq (\mathfrak{C}[t(\Sigma) \cup \left\{ !(x = y | x, y \in V) \right\}], \widehat{\Vdash})$, where $\Vdash$ is the smallest relation with the axioms of the equality theory:

$\Vdash !(x = x)$; $!(x = y) \Vdash !(y = x)$; $!(x = y) \otimes !(y = z) \Vdash !(x = z)$;

$f(x_0, \ldots, x_n) \otimes !(x_0 = y_0) \otimes \cdots \otimes !(x_n = y_n) \Vdash f(y_0, \ldots, y_n)$

Context-Free Constraint Grammars (CFCG).

Definition

A **context-free constraint grammar** (CFCG) is a tuple $(V, C, \vdash, \Sigma, P)$ where:

- V is a set of variables and $(C, \vdash)$ is a constraint-system over V;
- Σ is the signature of non-terminal symbols:
  $N \doteq \{f(y_1, \ldots, y_n) | f/n \in \Sigma$ and $y_1, \ldots, y_n \in V$ pair-wise distincts$\}$
- $P \subseteq N \times \overline{C}$ is the set of productions, with $\overline{C} \doteq \mathfrak{C}[C \cup N]$.
  Every production $(f(y_1, \ldots, y_n), u) \in P$ is such that $\mathrm{fv}(u) \subseteq \{y_1, \ldots, y_n\}$.
  A production is denoted: $f(y_1, \ldots, y_n) ::= u$.

Example:

$$h(X, Y) ::= \mathrm{edge}(X, Y)$$
$$h(X, Y) ::= \exists Z.\mathrm{edge}(X, Z) \otimes h(Z, Y)$$

Derivations for CFCGs.

Let $(V, C, \vdash, \Sigma, P)$ be a context-free constraint grammar. Let $\overline{C} \doteq \mathfrak{C}[C \cup N]$.

Definition

Let $\rightarrow \subseteq \overline{C}^2$ be the smallest relation satisfying the following rules:

$$\frac{f(y_1, \ldots, y_n) ::= u}{(f(y_1, \ldots, y_n))\sigma \rightarrow (u)\sigma} \quad \frac{u \rightarrow v}{\exists x. u \rightarrow \exists x v} \quad \frac{u \rightarrow v}{u \otimes w \rightarrow v \otimes w} \quad \frac{u \rightarrow v}{w \otimes u \rightarrow w \otimes v}$$

A derivation is a sequence $(c_i)_i$ with elements in $\overline{C}$ such that $c_0 \rightarrow c_1 \rightarrow \ldots$.

Example:

$$h(X, Y) ::= \text{edge}(X, Y)$$
$$h(X, Y) ::= \exists Z.\text{edge}(X, Z) \otimes h(Z, Y)$$

$$\begin{aligned} h(A, B) &\rightarrow \exists X_1.\text{edge}(A, X_1) \otimes h(X_1, B) \\ &\rightarrow \exists X_1 X_2.\text{edge}(A, X_1) \otimes \text{edge}(X_1, X_2) \otimes h(X_2, B) \\ &\rightarrow \exists X_1 X_2.\text{edge}(A, X_1) \otimes \text{edge}(X_1, X_2) \otimes \text{edge}(X_2, B) \end{aligned}$$

Languages accepted by CFCGs.

Let $G \doteq (V, C, \vdash, \Sigma, P)$ be a context-free constraint grammar. Let $\overline{C} \doteq \mathfrak{C}[C \cup N]$.

Definition

The **constraint language** accepted by a schema $c_0 \in \overline{C}$ and the grammar G is:

$$L_G(c_0) = \{c \in C | (\exists n)(\exists (c_i)_{0 \leq i \leq n})(c_0 \to \cdots \to c_n \wedge c_n \dashv\vdash c)\}$$

Example:

$$h(X, Y) ::= \text{edge}(X, Y)$$
$$h(X, Y) ::= \exists Z.\text{edge}(X, Z) \otimes h(Z, Y)$$

$$L_G(h(A, B)) = \left\{\text{hamiltonian paths from } A \text{ to } B\right\}$$

Link with Hyperedge Replacement Graph Grammars.

Claim

Over linear token systems, context-free constraint grammars have the same expressive power than hyperedge replacement graph grammars.

H. Ehrig, M. Nagl, and G. Rozenberg, editors. *Graph Grammars and Their Application to Computer Science*, Lect. Notes Comp. Sci. 153. Springer, 1983.

Theorem (K.-J. Lange and E. Welzl 87)

Hyperedge replacement graph language membership problem is NP-complete.

Hypergraphs.

A **labelled set** $\dot{E}$ is a tuple $(E, E^{\Sigma}, E^{\lambda})$, with E^{Σ} alphabet and $E^{\lambda} : E \to E^{\Sigma}$.

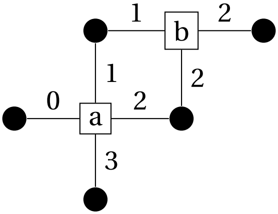
Definition

A **hypergraph** is a tuple $(V, \dot{E}, S, \to)$, where:

- V is the set of vertices; $\dot{E}$ is the labelled set of hyperedges;
- S is the alphabet of selectors;
- $\to \subseteq E \times S \times V$ is the incidence relation. $(e, s, v) \in \to$ is denoted $e \xrightarrow{s} v$.
  $\to$ is functional: $(\forall e\ s\ v_1\ v_2) e \xrightarrow{s} v_1 \wedge e \xrightarrow{s} v_2 \Rightarrow v_1 = v_2$

Vertices are represented as fat dots, hyperedges as boxes.

Let $\mathcal{H}(V, E, S)$ the set of hypergraphs over V, E, S.



Embedding of Hypergraphs into Constraints.

Let $H = (V, \dot{E}, S, \rightarrow)$ be a hypergraph.

- For all $e \in E$, let $\mathrm{dom}(e) \doteq \left\{ s \in S | (\exists v \in V)(e \xrightarrow{s} v) \right\}$.
- Let $n(e) \doteq |\mathrm{dom}(e)|$ and $r_e : \{1, \ldots, n(e) \hookrightarrow \mathrm{dom}(e)\}$.
- Let $\Sigma \doteq \left\{ f / n \in E^{\Sigma} \times \mathbb{N} | (\exists e \in E)(f = E^{\lambda}(e) \text{ and } n = n(e)) \right\}$.

Definition

The **constraint-embedding** of the hypergraph h is the following constraint of $\textsc{tok}(\Sigma)$:

$$[\![h]\!] \doteq (\exists v_1 \ldots v_n) \left(\bigotimes_{e \in E \wedge n = n(e) \wedge y_i = r_e(i)} [E^{\lambda}(e)](y_1, \ldots, y_n) \right)$$

Proposition

For every closed constraint $c \in \textsc{tok}(\Sigma)$, there exists a hypergraph h such that

$$[\![h]\!] \dashv\vdash c$$

Hypergraph Isomorphism.

Definition

Two hypergraphs $h_0 \doteq (V, (E, E^\Sigma, E^\lambda_0), S, \rightarrow_0)$ and $h_1 \doteq (V, (E, E^\Sigma, E^\lambda_1), S, \rightarrow_1)$ are **isomorphic**, denoted $h_0 \simeq h_1$, when there exist $\sigma_V : V \hookrightarrow V$ and $\sigma_E : E \hookrightarrow E$ such that:

- $(\forall e \in E)(E^\lambda_0(e) = E^\lambda_1(\sigma_E(e)))$;
- $(\forall e \in E)(\forall s \in S)(\forall v \in V)(e \xrightarrow{s}_0 v \Longleftrightarrow (\sigma_E(e) \xrightarrow{s}_1 \sigma_V(v))$.

Proposition

$$h_0 \simeq h1 \Longleftrightarrow [\![h_0]\!] \dashv\vdash [\![h_1]\!]$$

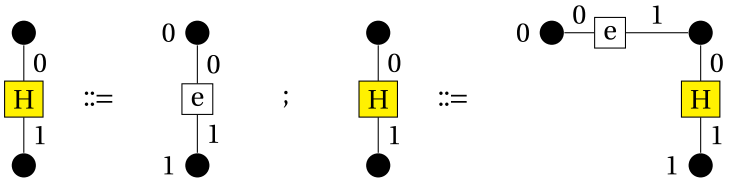
Hyperedge Replacement Graph Grammars (HRGs).

Let N, V, E, E^{Σ}, S be defined accordingly to the following definition.
We denote $\mathscr{H}^{+}(N, V, E, E^{\Sigma}, S)$ the set of **augmented hypergraphs**, that is hypergraphs whose edges are labelled with $\widehat{E}^{\Sigma}$ and vertices are in $\widehat{V}$.

Definition

A **h-edge replacement graph grammar** is a tuple $(N, V, E, E^{\Sigma}, S, P)$, with:

- N is the alphabet of non-terminals, such that $E^{\Sigma} \cap N = \emptyset$; $\widehat{E}^{\Sigma} \doteq E^{\Sigma} \uplus N$;
- V is the set of vertices;
- E is the set of hyperedges and E^{Σ} the set of ground labels;
- S is the alphabet of selectors, such that $V \cap S = \emptyset$; $\widehat{V} \doteq V \uplus S$;
- $P \subseteq N \times \mathfrak{P}(S) \times \mathscr{H}^{+}(N, V, E, E^{\Sigma}, S)$ is the set of productions.



Derivations for HRGs.

Let $(N, V, E, E^{\Sigma}, S, P)$ be a hyperedge replacement graph grammar.

Definition

Let $\rightarrow \subseteq \mathcal{H}^{+}(N, V, E, E^{\Sigma}, S)^2$ be the relation such that:
$(V_0, \dot{E}_0, S, \rightarrow_0) \rightarrow (V_1, \dot{E}_1, S, \rightarrow_1)$ iff there exist:

- a hyperedge $e \in E_0$; a production $(n, S_p, (V_p, \dot{E}_p, S, \rightarrow_p)) \in P$;
- $\sigma^E_{0\rightarrow 1} : E_0 \setminus \{e\} \hookrightarrow E_1$, $\sigma^V_{0\rightarrow 1} : V_0 \hookrightarrow V_1$, $\sigma^E_{p\rightarrow 1} : E_p \hookrightarrow E_1$, $\sigma^V_{p\rightarrow 1} : V_p \setminus S \hookrightarrow V_1$

such that:

- $E^{\lambda}_0(e) = n$; $\mathrm{im}(\sigma^E_{0\rightarrow 1}) \uplus \mathrm{im}(\sigma^E_{p\rightarrow 1}) = E_1$ and $\mathrm{im}(\sigma^V_{0\rightarrow 1}) \uplus \mathrm{im}(\sigma^V_{p\rightarrow 1}) = V_1$;
- $(\forall e_0 \in E_0 \setminus \{e\})(\forall e_p \in E_p)(\forall s\, s' \in S)(\forall v_0 \in V_0)(\forall v_p \in V_p \setminus S)$

$$E^{\lambda}_0(e_0) = E^{\lambda}_1(\sigma^E_{0\rightarrow 1}(e_0))$$
$$E^{\lambda}_p(e_p) = E^{\lambda}_1(\sigma^E_{p\rightarrow 1}(e_p))$$

$$e_0 \xrightarrow{s}_0 v_0 \Longleftrightarrow \sigma^E_{0\rightarrow 1}(e_0) \xrightarrow{s}_1 \sigma^V_{0\rightarrow 1}(v_0)$$
$$e_p \xrightarrow{s}_p v_p \Longleftrightarrow \sigma^E_{p\rightarrow 1}(e_p) \xrightarrow{s}_1 \sigma^V_{p\rightarrow 1}(v_p)$$
$$e \xrightarrow{s}_0 v_0 \wedge e_p \xrightarrow{s'}_p s \Longleftrightarrow \sigma^E_{p\rightarrow 1}(e_p) \xrightarrow{s'}_1 \sigma^V_{0\rightarrow 1}(v_0)$$

Embedding of HRGs into CFCGs.

Let $G \doteq (N, V, E, E^{\Sigma}, S, P)$ be a hyperedge replacement graph grammar.

1. $[\![\bullet]\!]$ can be extended to an embedding of augmented hypergraphs into scheme of $\overline{C}$, with $N \doteq \{f(y_1, \ldots y_n) | (\exists h)((f, \{y_1, \ldots, y_n\}, h) \in P)\}$;
2. therefore $[\![\bullet]\!]$ can be extended to an embedding of HRG productions into context-free constraint productions;
3. therefore $[\![\bullet]\!]$ can be extended to an embedding of HRG grammars into context-free constraint grammars.

Definition

The **hypergraph language** accepted by an augmented hypergraph $h_0 \in \mathscr{H}^+(N, V, E, E^{\Sigma}, S)$ and the grammar G is:

$$L_G(h_0) = \{h \in \mathscr{H}(V, E, S) | (\exists n)(\exists (h_i)_{0 \leq i \leq n})(h_0 \to \cdots \to h_n \wedge h \simeq h_n)\}$$

Proposition

For all $h_0 \in \mathscr{H}^+(N, V, E, E^{\Sigma}, S)$, we have $[\![L_G(h)]\!] = L_{[\![G]\!]}([\![h]\!])$.

With Equality Theory.

Lemma: for all constraint $c \in \text{TOK}_=(\Sigma)$, there exists a constraint $c' \in \text{TOK}(\Sigma) \subseteq \text{TOK}_=(\Sigma)$ such that $c \dashv\vdash c'$.

Proposition

For every closed constraint $c \in \text{TOK}_=(\Sigma)$*, there exists a hypergraph h such that*

$$[\![h]\!] \dashv\vdash c$$

Lemma: For a schema $c \in \overline{\text{TOK}_=(\Sigma)}$ and a grammar G over $\text{TOK}_=(\Sigma)$, there exists $c' \in \overline{\text{TOK}(\Sigma)}$ and a G' over $\text{TOK}(\Sigma)$ such that:

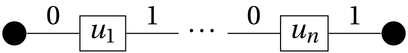
$$L_G(c) = \left\{h \in \text{TOK}_=(\Sigma) | \exists h' \in L_{G'}(c'), h \dashv\vdash h'\right\}$$

Proposition

Over $\text{TOK}_=(\Sigma)$*, context-free constraint language membership problem is NP-complete.*

Applications to Global Constraints.

Let A be an alphabet. Every $u = u_1 \cdots u_n \in A^*$ can be represented as a string-hypergraph:



Proposition (Habel and Kreowski 87)

Every context-free string language G can be recognized by a hyperedge replacement graph grammar $[\![G]\!]$.

Corollary: Every context-free string language G can be recognized by a context-free constraint grammar $[\![G]\!]$.

Proposition

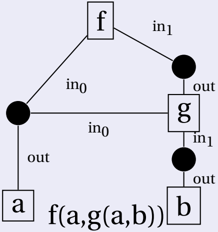
$$\text{CONTEXT-FREE}(x_1, \ldots, x_n, G) \doteq (\exists)\, [\![G]\!]\, (x_1, \ldots, x_n)$$

Limitations.

Definition

A **jungle** is a hypergraph $(V, \dot{E}, \{\text{in}_k, \text{out}\}, \rightarrow)$ such that:

- $(\forall v \in V)\left(\left|\left\{e \in E | e \overset{\text{out}}{\rightarrow} v\right\}\right| \leq 1\right)$;
- $\rightarrow$ is acyclic.



F. Raiser and T. Frühwirth, *Towards Term Rewriting Systems in Constraint Handling Rules – Coming to terms with jungles*, Proceedings of the Fifth Workshop on Constraint Handling Rules, 2008.

"**Corollary 2:** CHR folding is terminating and confluent."

Only true if the query is a jungle!

Proposition

$\{h \in \mathcal{H}(V, E, S) | h \textit{ is a jungle}\}$ *is not a hyperedge replacement graph language.*

Domain-Sensitive Constraint Grammars. (1/2)

A **two-sorted signature** Σ is a subset of $F \times \mathbb{N} \times \mathbb{N}$, where F is a countable symbol set. Items $(f, n, m) \in \Sigma$ are denoted $f/(n, m)$.
Let V and W be two set of variables and L a set of labels.
The **set constraint system** is the smallest *classical* constraint system $(S, \vdash_S)$ such that:

$$\frac{w \in W}{w = \emptyset \in S} \quad \frac{v \in V \quad w \in W}{w = \{v\} \in S} \quad \frac{v \in V \quad w \in W}{(v \in w) \in S} \quad \frac{w \in W \quad l \in L}{w = \mathrm{fv}\, l \in S}$$

$$\frac{w_1 \in W \quad w_2 \in W \quad w_3 \in W}{w_1 = w_2 \; op \; w_3 \in S} \; op \in \{\cap, \cup, \setminus\}$$

with $w = \{v\} \vdash_S v \in w \quad w = \emptyset \vdash_S v \notin w$
and $v \in w_1 \; op_1 \; v \in w_2 \vdash_S v \in w_1 \; op_2 \; w_2$ with
$(op_1, op_2) \in \{(\wedge, \cap); (\vee, \cup); (\wedge\neg, \cap)\}\ldots$

Domain-Sensitive Constraint Grammars. (2/2)

Definition

A **domain-sensitive constraint grammar** (DSCG) is a tuple $(V, W, L, C, \vdash, \Sigma, P)$ where:

- V, W are sets of variables, L a set of labels;
- $(C, \vdash)$ is a constraint-system over V;
- Σ is the two-sorted signature of non-terminal symbols:
  $$N \doteq \left\{ l : f(y_1, \ldots, y_n, Y_1, \ldots, Y_m) \mid l \in L, f/(n,m) \in \Sigma \text{ and } \begin{array}{c} y_1, \ldots, y_n \in V \\ Y_1, \ldots, Y_m \in W \end{array} \text{ p.-w. dist.} \right\}$$
- $P \subseteq N \times \overline{C} \times S$ is the set of productions, with $\overline{C} \doteq \mathfrak{C}[C \cup N]$.
  A production is denoted: $f(y_1, \ldots, y_n, Y_1, \ldots, Y_m) ::= u | g$.

Applications of DSCGs.

Jungles Let Σ be a signature. For all $f/n \in \Sigma$:

$$j(r) ::= \mathrm{eq}(r, f(x_1, \ldots, x_n)) \otimes l_1 : j(x_1) \otimes \cdots \otimes l_n : j(x_n) | r \notin \mathrm{fv}(l_1) \cup \cdots \cup \mathrm{fv}(l_n)$$

all-different

$$a(S) ::= x_1 \neq x_2 | x_1 \in S \wedge x_2 \in S \setminus \{x_1\}$$

$$\textsc{all-different}(x_1, \ldots x_n) \doteq (\forall) a(\{x_1, \ldots, x_n\})$$

Conclusions & Perspectives.

Objective

- Describing set of constraints;
- for observational confluence;
- for defining global constraints.

Contributions

- Definition of constraint grammar;
- as expressive as hyperedge-replacement graph grammar;
- much more easy to define;
- more general.

Future work

- Decidability and complexity analysis for other constraint systems.